

Rates Achievable on a Fiber-Optical Split-Step Fourier Channel

Kamran Keykhosravi, *Student Member, IEEE*,

Erik Agrell, *Senior Member, IEEE*, Giuseppe Durisi, *Senior Member, IEEE*

Abstract

A lower bound on the capacity of the split-step Fourier channel is derived. The channel under study is a concatenation of smaller segments, within which three operations are performed on the signal, namely, nonlinearity, linearity, and noise addition. Simulation results indicate that for a fixed number of segments, our lower bound saturates in the high-power regime and that the larger the number of segments is, the higher is the saturation point. We also obtain an alternative lower bound, which is less tight but has a simple closed-form expression. This bound allows us to conclude that the saturation point grows unbounded with the number of segments. Specifically, it grows as $c + (1/2) \log(K)$, where K is the number of segments and c is a constant. The connection between our channel model and the nonlinear Schrödinger equation is discussed.

Index Terms

This work was supported by the Swedish Research Council (VR) under Grant 2013-5271. The simulations were executed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at C3SE. This work was presented in part at the Munich Workshop on Information Theory of Optical Fiber, 2015 (not published as a paper).

The authors are with the Department of Signals and Systems, Chalmers University of Technology, Gothenburg 41296, Sweden (e-mail: kamrank@chalmers.se; durisi@chalmers.se; agrell@chalmers.se).

I. INTRODUCTION

Finding the capacity of the nonlinear and dispersive optical channel is a formidable task, so much so that not only the capacity has not been established, but also a large gap between the known upper and lower bounds exists. While all known lower bounds either saturate or fall to zero in the high-power regime, the only available upper bound [1], [2] grows logarithmic with the power, i.e., it behaves as $\log(1 + \text{SNR})$, where SNR is the signal-to-noise ratio. Neglecting dispersion, the channel capacity can be calculated [3], [4]. In this case, its asymptotic behavior in the high-power regime is $(1/2) \log(\text{SNR}) - 1/2$.

Many lower bounds have been proposed on the capacity of fiber-optical channels (see for example [5]–[9]), most of which fall to zero at high powers. Consequently, it was widely believed that the capacity would diminish at high powers. Recent works disprove this belief [10], [11]. However, to the best of our knowledge, no lower bound has been established as yet that grows unbounded with power.

The nonlinear Schrödinger equation (NLSE) models the fiber-optical channel excellently. However, it is not suitable for information theory analyses since its input and output are continuous-time waveforms. The split-step Fourier (SSF) method is a standard method to simulate the NLSE and has been validated by many experiments. The SSF method approximates the NLSE by discretizing it in time and space (by splitting the fiber channel into multiple segments). Thus, its input and output are complex vectors. Moreover, the output vector can be obtained by recursive computations over the many channel segments. This method has been used in [1] to establish an upper bound on the capacity of the fiber-optical channel.

The accuracy of the SSF method depends on the step size in the spatial domain as well as the sampling interval in the temporal domain. When the number of segments goes to infinity,

the SSF method approximates the NLSE accurately [12, Sec. 4.2.1]. The error caused by using a finite number of segments depends on the input power: to maintain a desired accuracy level as the input power grows, the number of segments needs to increase (or, equivalently, the step size needs to decrease) with the input power. To the best of our knowledge, no closed-form expression is available for the SSF method error as a function of the input power and the step size. However, results based on simulations and on approximations of the error are available (see for example [13]).

Because of the nonlinearity, the bandwidth of the input signal broadens as it propagates along the fiber. In practice, this effect is taken into account in the SSF model by oversampling, i.e., by sampling faster than the Nyquist rate. In this paper, we ignore the effects of spectrum broadening, which is left to future studies, and consider an SSF model with sampling performed at the Nyquist rate.

Contributions: We derive a lower bound on the capacity of the SSF fiber-optical channel with K segments, using as input vector distribution a multivariate Gaussian with independent and identically distributed (i.i.d.) components. We present a lower bound that can be evaluated by calculating through Monte Carlo simulations the expectation of a function of input and noise vectors. The simulation results indicate that the bound saturates at high powers and that the saturation point increases with the number of segments. We further lower-bound the aforementioned bound by a closed-form expression, which reveals that the saturation point increases by 0.5 bit whenever the number of segments is doubled. This unbounded increase of our capacity lower bound casts doubt on the existence of an optimal input power, going beyond which is idle or even detrimental to the optical system performance.

The outline of this paper is as follows. In Section II, we describe the continuous-time and the discrete-time channel models. In Section III, we state our simulation-based as well as closed-form lower bounds. The simulation results are provided in Section IV. The

first Appendix contains some preliminary results, which come into use in the subsequent Appendices, where our theorems are proved.

Notation: We use boldface letters to denote random quantities. Vectors, which are columns by default, are identified by underlined letters, whereas matrices are denoted by upper-case *sans-serif* letters (e.g., \mathbf{A}). The identity matrix of size $L \times L$ is denoted by \mathbf{I}_L . The i th element of a vector is indicated by the subscript i . For a complex number x , we denote its real part, imaginary part, absolute value, and phase by $\mathcal{R}(x)$, $\mathcal{I}(x)$, $|x|$, and $\angle x$, respectively. The Euclidean norm of $\underline{x} \in \mathbb{C}^L$ is denoted by $\|\underline{x}\|$; also, we let $\llbracket \underline{x} \rrbracket^2$ be the vector whose i th element is $|x_i|^2$. We use $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^\dagger$ to indicate the transpose, complex conjugate, and conjugate transpose operators, respectively. Furthermore, $\mathbb{E}_{\mathbf{x}}[\cdot]$ denotes expectation with respect to the random variable \mathbf{x} . The subscript will be omitted if obvious from the context. We use $\text{Tr}[\cdot]$ to indicate the trace operator. We let $\mathcal{CN}(\underline{0}, \mathbf{K})$ be the multivariate complex-valued circularly symmetric Gaussian distribution with covariance matrix \mathbf{K} . We use $\text{unif}(a, b)$ to denote a uniform distribution over the interval $[a, b]$. The indicator function is denoted by $\mathbf{1}(\cdot)$. All logarithms are in base two.

II. CHANNEL MODEL

Optical fiber systems employ optical amplification to compensate for losses in the fiber at the expense of an increased noise level. Two amplification principles exist, namely, lumped and distributed amplification. Lumped amplification makes use of several amplifiers along the fiber. Distributed amplifications compensates for the energy loss continuously, so that the signal energy level remains roughly constant throughout the propagation. Throughout the paper, we consider the ideal distributed-amplification case, that is, the signal power is assumed to be constant throughout the propagation.

A. Continuous-Time Model

For distributed amplification, the generalized NLSE captures the effect of nonlinearity, dispersion, and noise along the fiber [14]. It is a nonlinear partial differential equation and can be written as

$$\frac{\partial \mathbf{a}}{\partial z} + j\frac{\beta_2}{2}\frac{\partial^2 \mathbf{a}}{\partial t^2} - j\gamma|\mathbf{a}|^2\mathbf{a} = \mathbf{n}. \quad (1)$$

Here, γ and β_2 are the nonlinear coefficient and the group-velocity dispersion parameter, respectively. The variable $\mathbf{a} = \mathbf{a}(z, t)$ indicates the complex envelope of the optical field in location z and at time t . Furthermore, $\mathbf{n}(z, t)$, which is a complex-valued zero-mean Gaussian process, models the amplification noise. This process is spatially white and its power spectral density $S_{\mathbf{n}}(f)$ is given by $S_{\mathbf{n}}(f) = N_{ASE}/Z$, where Z is the fiber length and N_{ASE} is the noise power spectral density at the receiver. Equation (1), which unfortunately admits no analytical solution, can be regarded as a continuous-time channel model for a fiber-optical link, with input waveform $\mathbf{a}(0, t)$ and output waveform $\mathbf{a}(Z, t)$.

B. Discrete-Time Model

We move from continuous time to discrete time by sampling the input signal every Δ_t seconds. Through this sampling technique, we map an input signal of duration $T - \Delta_t$ seconds into a complex vector $\underline{\mathbf{a}}_0$ of dimension $L = T/\Delta_t$. Similarly, at the receiver, we sample the output signal and obtain a complex vector.

The map between input and output vectors can be approximated by using the SSF method, which approximates the fiber-optical channel by a cascade of K segments of length $\Delta_z = Z/K$. For a fixed fiber length Z , the SSF method gets precise as K goes to infinity (or, equivalently, as Δ_z goes to zero). For $1 \leq k \leq K$, we denote the output vector of the k th segment by $\underline{\mathbf{a}}_k = [\mathbf{a}_{k,0}, \dots, \mathbf{a}_{k,L-1}]^T$ (see Fig. 1); also, we use $\underline{\mathbf{a}}_0$ to denote the input vector. The relations between the discrete-time and the continuous-time channel inputs and outputs are $\mathbf{a}_{0,l} = \mathbf{a}(0, l\Delta_t)$ and $\mathbf{a}_{K,l} = \mathbf{a}(Z, l\Delta_t)$, for $0 \leq l \leq L - 1$.

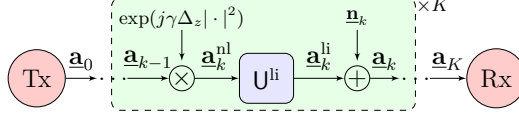


Fig. 1: The channel model based on the split-step Fourier method.

The output of each segment is computed by separating the linear and the nonlinear operations as illustrated in Fig. 1. Specifically, the evolution from $\underline{\mathbf{a}}_{k-1}$ to $\underline{\mathbf{a}}_k$ ($1 \leq k \leq K$) involves the following three steps:

- 1) *Nonlinear step* (the Kerr effect): For $0 \leq l \leq L-1$,

$$\mathbf{a}_{k,l}^{\text{nl}} = \mathbf{a}_{k-1,l} e^{j\gamma|\mathbf{a}_{k-1,l}|^2\Delta z}. \quad (2)$$

- 2) *Linear step* (chromatic dispersion):

$$\underline{\mathbf{a}}_k^{\text{li}} = \mathbf{U}^{\text{li}} \underline{\mathbf{a}}_k^{\text{nl}}. \quad (3)$$

Here, \mathbf{U}^{li} is a unitary matrix defined by

$$\mathbf{U}^{\text{li}} = \mathbf{F}^\dagger \mathbf{D}^{\text{li}} \mathbf{F} \quad (4)$$

where \mathbf{F} is the discrete Fourier transform (DFT) operator with entries

$$f_{l,m} = \frac{1}{\sqrt{L}} e^{-j2\pi lm/L}, \quad 0 \leq l, m \leq L-1. \quad (5)$$

Furthermore, $\mathbf{D}^{\text{li}} = \text{diag}\{d_0, \dots, d_{L-1}\}$ is a diagonal matrix with entries

$$d_l = e^{j\Delta z f^{\text{li}}(l)}, \quad l = 0, \dots, L-1 \quad (6)$$

where

$$f^{\text{li}}(l) = \frac{\beta_2}{2} \left(\frac{2\pi}{T} \right)^2 \left(\frac{L}{2} - \left| \frac{L}{2} - l \right| \right)^2. \quad (7)$$

For efficient implementation, (4) is usually computed using the fast Fourier transform.

3) *Noise addition:*

$$\underline{\mathbf{a}}_k = \underline{\mathbf{a}}_k^{\text{li}} + \underline{\mathbf{n}}_k \quad (8)$$

where $\underline{\mathbf{n}}_k \sim \mathcal{CN}(\underline{0}, \sigma_n^2 \mathbf{I}_L)$ with

$$\sigma_n^2 = \frac{P_n}{K} \quad (9)$$

where P_n is the per-sample noise variance, which can be calculated using the parameters of the noise generated by the inline amplifiers as [15]

$$P_n = N_{ASE} B_n \quad (10)$$

$$= h\nu Z \alpha n_{sp} B_n. \quad (11)$$

Here, $h\nu$ is the optical photon energy, α is the attenuation parameter, n_{sp} is the spontaneous emission factor, and B_n is the receiver filter bandwidth. Due to nonlinear effects, each signal frequency component interacts with all possible noise frequency components. However, this interaction becomes weaker as the frequency gap between these two components increases [16]. Here, we assume that the bandwidth of the receive filter is much greater than that of the signal (i.e., $B_n \Delta_t \gg 1$) so that the influence of the interaction between the out-of-band noise and the signal can be neglected.

Using K times the three steps listed above, we obtain a probabilistic channel law that maps the input vector $\underline{\mathbf{a}}_0$ into the output vector $\underline{\mathbf{a}}_K$. We shall refer to this law as the *SSF channel* with length Z and number of segments K . Assuming that the SSF channel is block-memoryless across blocks of length T seconds, we can write its capacity (in bits per channel use) as

$$\mathcal{C}_K = \frac{1}{L} \sup I(\underline{\mathbf{a}}_K; \underline{\mathbf{a}}_0). \quad (12)$$

Here, the supremum is over all probability distributions on the input random vector $\underline{\mathbf{a}}_0$ that satisfy the power constraint $\mathbb{E}[\|\underline{\mathbf{a}}_0\|^2] \leq LP$, where P is the input power.

The only known *upper* bound on the capacity (12) of the SSF channel is [1], [2]

$$\mathcal{C}_K \leq \log(1 + P/P_n). \quad (13)$$

This bound is valid for every K . In contrast, a multitude of *lower* bounds have been proposed. Most, if not all, such bounds use various mismatched decoding approaches, where nonlinear distortion is treated as noise at the receiver [5]–[9].

III. LOWER BOUNDS ON THE CAPACITY OF THE SSF CHANNEL

In this section, we propose one simulation-based as well as two closed-form lower bounds on the capacity of the SSF channel, given in Theorems 1–3. First, we lower-bound the capacity by a function that can be evaluated through Monte Carlo simulation. Second, we provide a lower bound on this function by a closed-form expression to analyze our simulation results at high power. Third, we replace our second bound by an explicit function of the input power and K (the number of segments) at the expense of tightness. We evaluate our first bound through simulations (see Section IV), which indicate that this bound saturates in the high-power regime. Our second bound, although loose at low powers, can be used to approximate the asymptotes of the simulation results. Finally, we use our third bound to show that these asymptotes go to infinity as K grows large. We use the following two lemmas to establish our first lower bound in Theorem 1.

Lemma 1. If the input vector distribution of the SSF channel is i.i.d. Gaussian, i.e., $\mathbf{a}_0 \sim \mathcal{CN}(\mathbf{0}, P\mathbf{I}_L)$, then

$$\mathbf{a}_K \sim \mathcal{CN}(\mathbf{0}, (P + P_n)\mathbf{I}_L). \quad (14)$$

Proof: See Appendix B. ■

Lemma 2. Let (\mathbf{x}, \mathbf{y}) be a pair of L -dimensional proper complex random vectors, distributed according to an arbitrary joint probability density function. The conditional entropy $h(\mathbf{y} | \mathbf{x})$

is bounded as

$$h(\underline{\mathbf{y}}|\underline{\mathbf{x}}) \leq \frac{L}{2} \log \left(2\pi^3 e^{\frac{\kappa(\underline{\mathbf{y}}|\underline{\mathbf{x}})}{L}} \right) \quad (15)$$

where

$$\kappa(\underline{\mathbf{y}}|\underline{\mathbf{x}}) = \sum_{i=0}^{L-1} \mathbb{E} \left[|\mathbf{y}_i|^4 \right] - \sum_{i=0}^{L-1} \mathbb{E}_{\underline{\mathbf{x}}} \left[\mathbb{E}_{\underline{\mathbf{y}}} \left[|\mathbf{y}_i|^2 | \underline{\mathbf{x}} \right]^2 \right]. \quad (16)$$

Proof: See Appendix C. ■

Based on these two lemmas, we can now formulate our first lower bound as follows.

Theorem 1. Let $\underline{\mathbf{a}}_0 \sim \mathcal{CN}(\underline{0}, PI_L)$. The capacity (12) of the SSF channel is lower-bounded as

$$\mathcal{C}_K \geq L_K^{(1)} = \frac{1}{2} \log \left(\frac{e}{2\pi} \cdot \frac{(P + P_n)^2}{2(P + P_n)^2 - \mathcal{E}} \right) \quad (17)$$

where

$$\mathcal{E} = \frac{1}{L} \sum_{i=0}^{L-1} \mathbb{E}_{\underline{\mathbf{a}}_0} \left[\mathbb{E} \left[|\mathbf{a}_{K,i}|^2 | \underline{\mathbf{a}}_0 \right]^2 \right]. \quad (18)$$

Proof: By (12), we have that $\mathcal{C}_K \geq I(\underline{\mathbf{a}}_K; \underline{\mathbf{a}}_0) / L$, where $\underline{\mathbf{a}}_0 \sim \mathcal{CN}(\underline{0}, PI_L)$. Next, we decompose $I(\underline{\mathbf{a}}_K; \underline{\mathbf{a}}_0)$ as

$$I(\underline{\mathbf{a}}_K; \underline{\mathbf{a}}_0) = h(\underline{\mathbf{a}}_K) - h(\underline{\mathbf{a}}_K | \underline{\mathbf{a}}_0). \quad (19)$$

For our choice of input distribution, Lemma 1 yields that $\underline{\mathbf{a}}_K \sim \mathcal{CN}(\underline{0}, (P + P_n) I_L)$. Hence, we conclude that [17, Thm. 2]

$$h(\underline{\mathbf{a}}_K) = L \log(\pi e (P + P_n)). \quad (20)$$

Furthermore, it follows from Lemma 2 that

$$h(\underline{\mathbf{a}}_K | \underline{\mathbf{a}}_0) \leq \frac{L}{2} \log \left(2\pi^3 e^{\frac{\kappa(\underline{\mathbf{a}}_K | \underline{\mathbf{a}}_0)}{L}} \right), \quad (21)$$

where, by definition (see (16)),

$$\kappa(\underline{\mathbf{a}}_K | \underline{\mathbf{a}}_0) = \sum_{i=0}^{L-1} \mathbb{E} \left[|\mathbf{a}_{K,i}|^4 \right] - \sum_{i=0}^{L-1} \mathbb{E}_{\underline{\mathbf{a}}_0} \left[\mathbb{E}_{\underline{\mathbf{a}}_K} \left[|\mathbf{a}_{K,i}|^2 | \underline{\mathbf{a}}_0 \right]^2 \right] \quad (22)$$

$$= (2(P + P_n)^2 - \mathcal{E}) L. \quad (23)$$

Here, in the last step we used that $\mathbf{a}_{K,i} \sim \mathcal{CN}(0, P + P_n)$. Substituting (23) into (21) and then (20) and (21) into (19), we obtain (17). \blacksquare

In the absence of a closed-form expression, (17) can be calculated by evaluating \mathcal{E} through Monte Carlo simulation (see Section IV). Further lower-bounding $L_K^{(1)}$ in (17), one can obtain an expression that can be evaluated in closed form, as shown in the next theorem.

Theorem 2. For every $0 \leq m \leq L - 1$, let

$$\alpha_{0,m} = \frac{1}{L} \sum_{l=0}^{L-1} e^{j\Delta_z f^{\text{li}}(l)} e^{-j2\pi lm/L} \quad (24)$$

$$\alpha = \left(\sum_{m=0}^{L-1} |\alpha_{0,m}| \right)^2 - \sum_{m=0}^{L-1} |\alpha_{0,m}|^2 \quad (25)$$

$$D = \alpha K \left(\frac{K^3}{2e\gamma^2 P_n Z^2} + \frac{K}{eZ\gamma} + \frac{P_n}{2eK} \right) \quad (26)$$

and

$$\zeta(P, K) = |\alpha_{0,0}|^{4(K-1)} \exp\left(-\frac{D}{P|\alpha_{0,0}|^{2(K-1)}}\right) \quad (27)$$

Then $L_K^{(1)}$ in (17) can be lower-bounded as

$$L_K^{(1)} \geq L_K^{(2)} = \frac{1}{2} \log \left(\frac{e}{4\pi} \cdot \frac{(P + P_n)^2}{(P + P_n)^2 - \zeta(P, K) P^2} \right). \quad (28)$$

Proof: By comparing (17) and (28), we see that, to prove Theorem 2, it is sufficient to show that $\mathcal{E} \geq 2P^2\zeta(P, K)$, where \mathcal{E} is defined in (18). In Appendix D, we prove that for every $0 \leq r \leq L - 1$

$$\mathbb{E}_{\mathbf{a}_0} \left[\left(\mathbb{E}_{\mathbf{a}_K} \left[|\mathbf{a}_{K,r}|^2 \mid \mathbf{a}_0 \right] \right)^2 \right] \geq 2P^2\zeta(P, K) \quad (29)$$

from which the desired result follows. \blacksquare

For any given K , (28) yields

$$\lim_{P \rightarrow \infty} L_K^{(2)} = -\frac{1}{2} \log \left(\frac{4\pi}{e} \left(1 - |\alpha_{0,0}|^{4(K-1)} \right) \right) = \tilde{L}_K^{(2)}. \quad (30)$$

Note that $|\alpha_{0,0}| \leq 1$ which follows by triangle inequality, so the expression on the RHS of (30) is well defined. The limit in (30) reveals that $L_K^{(2)}$ approaches the constant $\tilde{L}_K^{(2)}$ as $P \rightarrow \infty$, and the value of this constant depends on K . We are interested in how $L_K^{(2)}$ and its asymptote limit $\tilde{L}_K^{(2)}$ behave as K grows large. This behavior will be illustrated numerically in Section IV, where we use $\tilde{L}_K^{(2)}$ to provide a lower bound on the high-power asymptote of $L_K^{(1)}$. In Theorem 3 we further lower-bound $L_K^{(2)}$ to obtain a less tight but simpler bound on \mathcal{C}_K , which reveals its dependence on K in the asymptotic limit $P \rightarrow \infty$.

Theorem 3. Let

$$C_1 = \frac{Z^2}{L} \sum_{l=0}^{L-1} \left(f^{\text{li}}(l) \right)^2, \quad (31)$$

$$C_2 = \sqrt{6}Z \sum_{l=0}^{L-1} f^{\text{li}}(l), \quad (32)$$

and

$$G = C_2 \left(\frac{C_2}{4K} + 1 \right) \left(\frac{K^3}{2e\gamma^2 P_n Z^2} + \frac{K}{eZ\gamma} + \frac{P_n}{2eK} \right). \quad (33)$$

Then, for every $K > \max \left\{ \frac{\beta_2 Z \pi^2}{2\sqrt{2}\Delta_t^2}, \sqrt{C_1} \right\}$, the bound $L_K^{(2)}$ in (28) is further lower-bounded as

$$L_K^{(2)} \geq L_K^{(3)} = \frac{1}{2} \log \left(\frac{e}{4\pi} \cdot \frac{(P + P_n)^2}{2C_1 P^2 / K + 2PP_n + PG + P_n^2} \right). \quad (34)$$

Proof: See Appendix E. ■

We use $L_K^{(3)}$ to lower-bound the asymptotic behavior of $L_K^{(1)}$ and $L_K^{(2)}$ as $P \rightarrow \infty$. Since we are interested in these asymptotes as K grows large, the condition on K mentioned in Theorem 3 is not restrictive. We obtain from (34) that

$$\lim_{P \rightarrow \infty} L_K^{(3)} = \frac{1}{2} \log \left(\frac{eK}{8\pi C_1} \right) = \tilde{L}_K^{(3)}. \quad (35)$$

To be more specific, it follows from (34) that if K grows with P so that $P/K^3 \rightarrow \infty$, then the lower bound $L_K^{(3)}$ goes to infinity. As mentioned in Section I, the SSF method with a fixed number of segments is a valid approximation of NLSE up to a certain power. Our lower

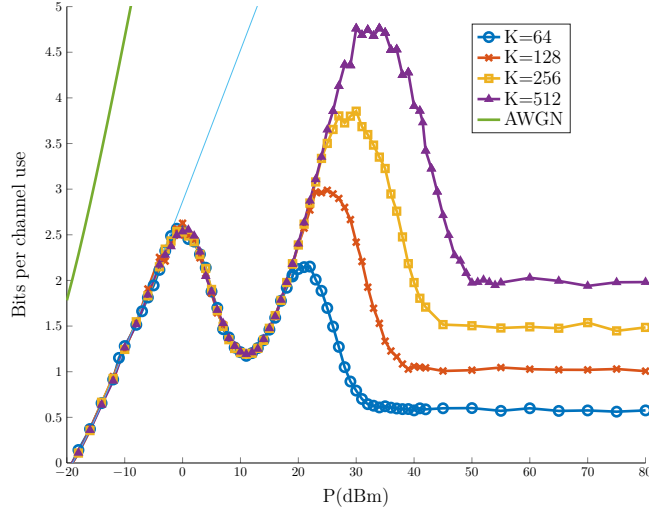


Fig. 2: Simulation results for the lower bound $L_K^{(1)}$ in (17), for a different numbers of channel segments K . The AWGN upper bound in (13) and the low-power approximation $L_{LP}^{(1)}$ in (36) are also shown.

bound $L_K^{(3)}$ indicates that if we increase the number of segments K with power P such that $P/K^3 \rightarrow \infty$ as $P \rightarrow \infty$, the capacity grows unboundedly.

IV. NUMERICAL EXAMPLE

In this section, we present and analyze the results obtained by evaluating $L_K^{(1)}$ in (17) through Monte Carlo simulation. After stating the channel parameters, we analyze our simulation results in low, moderate, and high power regimes separately. Finally, we draw conclusions based on our analytical and numerical results.

We consider a single-mode fiber link with parameters given in Table I. The per-sample noise variance can be calculated through (11) and is equal to $P_n = 4.1 \mu\text{W}$. Four different values of channel segments K are considered. They correspond to segment lengths Δ_z of (approximately) 13.3, 6.6, 3.3, and 1.7 km.

In Fig. 2, $L_K^{(1)}$ is numerically evaluated by Monte Carlo simulation. For every K and P , 200 independent realizations of \mathbf{a}_0 , with length $L = 2000$, were generated, and for each of

TABLE I: Channel parameters used in the example.

Parameter	Symbol	Value
Fiber length	Z	850 km
Attenuation	α	0.2 dB/km
Dispersion	β_2	$-21.7 \text{ ps}^2/\text{km}$
Nonlinearity	γ	1.27 (Wkm)^{-1}
Symbol time	Δ_t	100 ps
Optical photon energy	$h\nu$	$1.3 \cdot 10^{-19} \text{ J}$
Spontaneous emission factor	n_{sp}	4
Filter bandwidth	B_n	200 GHz

these realizations, 1000 independent realizations of \mathbf{a}_K were generated using (2)–(8).

As can be seen in Fig. 2, in the low-power regime the evaluation of $L_K^{(1)}$ results in the same lower bound for all the four considered values of K . This is because at low power, the SSF method models the NLSE accurately for all values of K considered here. Since in low-power regime the nonlinearity can be neglected, it is possible to obtain a closed-form accurate approximation of $L_K^{(1)}$ in this regime. Specifically, by setting $\gamma = 0$, the SSF channel turns into the linear channel $\mathbf{a}_K = (\mathbf{U}^{\text{li}})^K \mathbf{a}_0 + \mathbf{n}$, where $\mathbf{n} \sim \mathcal{CN}(0, P_n \mathbf{I}_L)$. Noting that $(\mathbf{U}^{\text{li}})^K \mathbf{a}_0 \sim \mathcal{CN}(0, P \mathbf{I}_L)$, one can evaluate \mathcal{E} for this channel in closed form and obtain

$$L_{\text{LP}}^{(1)} = \frac{1}{2} \log \left(\frac{e}{2\pi} \left(1 + \frac{P^2}{(2P + P_n) P_n} \right) \right). \quad (36)$$

This approximation, which is plotted in Fig. 2, is accurate for values of power P less than 0 dBm.

At moderate power levels, our bound shows a peak at approximately 0 dBm. We next provide an intuitive discussion to explain why our bound decreases in the interval [0 dBm , 10 dBm].

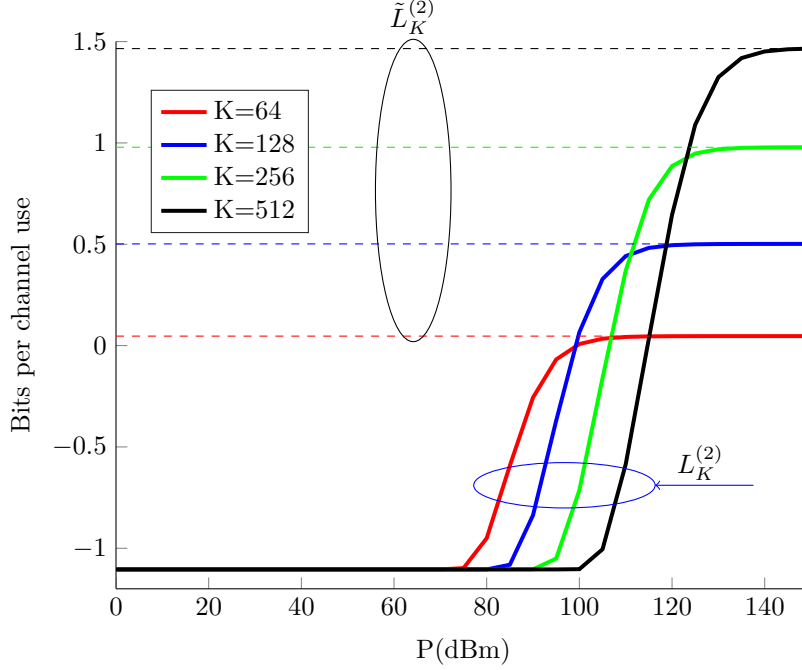


Fig. 3: Numerical evaluation of the lower bound $L_K^{(2)}$ in (28), for a different number of channel segments K . The high-power asymptote $\tilde{L}_K^{(2)}$ in (30) is also illustrated by horizontal dashed lines.

At moderate power levels, the effects of the nonlinearity become substantial. The interaction between the nonlinearity and the noise changes the phase of the signal randomly during propagation. This phase noise leads to amplitude noise when the chromatic dispersion is applied. Next, we show by an example that having amplitude randomness at the receiver causes an increase of $\kappa(\mathbf{a}_K|\mathbf{a}_0)$ and hence a decrease of $L_K^{(1)}$. Define a random vector $\tilde{\mathbf{a}}_K$ satisfying $|\tilde{\mathbf{a}}_{K,i}|^2 = |\mathbf{a}_{K,i}|^2 + \mathbf{n}_{p,i}$ for $i = 0, \dots, L-1$, where \mathbf{n}_p is a signal-independent zero-mean noise with covariance matrix $\tilde{P}_n \mathbf{I}_L$. Using the definition of $\kappa(\cdot|\cdot)$ in (16), we have

$$\kappa(\tilde{\mathbf{a}}_K|\mathbf{a}_0) = \kappa(\mathbf{a}_K|\mathbf{a}_0) + L\tilde{P}_n. \quad (37)$$

As shown in Fig. 2, our bound starts increasing again roughly at 10 dBm.

In the high-power regime, as can be seen in Fig. 2, $L_K^{(1)}$ becomes sensitive to K . This is due to the fact that for a fixed K , the SSF method is accurate up to a certain power. In

other words, as far as the calculation of $L_K^{(1)}$ is concerned, the SSF method with $K = 64$, 128, and 256 segments accurately models the continuous channel only up to the power levels of 19, 22, and 25 dBm, respectively.

As it is evident in Fig. 2, $L_K^{(1)}$ eventually saturates at high power levels, and the saturation point increases with K . We use our asymptotic closed-form bound $\tilde{L}_K^{(2)}$ in (30) to approximate these asymptotes. In Fig. 3, the lower bound $L_K^{(2)}$ is evaluated as a function of power for different numbers of channel segments K ; furthermore, the asymptote of this lower bound, $\tilde{L}_K^{(2)}$, is also shown via horizontal dashed lines. Comparing the results in Fig. 2 and Fig. 3, one observes that approximations made by $\tilde{L}_K^{(2)}$ (which are 0.05, 0.5, 0.98, and 1.47) are different from the simulation asymptotes (which are 0.58, 1, 1.49, and 1.98) by a constant value of approximately 0.5 bits, which becomes negligible as K becomes large. The asymptotic lower bound $\tilde{L}_K^{(3)}$ (35) suggests that the saturation point increases by 0.5 bits if the value of K doubles, which is evident in Figs. 2 and 3.

To summarize, we evaluated $L_K^{(1)}$ through simulation and observed that this lower bound increases with K . Next, we used our analytical bound $\tilde{L}_K^{(3)}$ to show that this trend will sustain for large values of K . Therefore, one may conclude that, as long as the effects of spectrum broadening are neglected, the capacity of the NLSE channel goes to infinity with power.

V. CONCLUSION

We presented a lower bound on the capacity of the split-step Fourier method channel, which can be evaluated by calculating a double expectation using Monte Carlo simulations. Doing so, we evaluated this bound for different numbers of channel segments K , and different transmit power levels. Simulation results indicated that for a fixed K , the lower bound saturates at high power and the saturation point increases with K . To study the asymptotic behavior of this bound, we further lower-bounded it by two closed-form expressions. Our

analytical results prove that with appropriate choices of power P and number of segments K , the capacity of the SSF channel can be made arbitrarily large. Using our analytical bounds, we proved that the saturation point increases to infinity as we increase the number of channel segments. Specifically, we showed that the asymptotes of our bound increase by 0.5 bit if the number of segments is doubled. Our numerical and analytical results suggest that as long as the effect of spectrum broadening is ignored, the capacity of the fiber-optical channel described by the NLSE goes to infinity with power.

APPENDIX A

PRELIMINARIES

A.I. Maximum Entropy

Among all real random vectors $\underline{\mathbf{x}}$ with a fixed nonsingular correlation matrix $\mathbf{R}(\underline{\mathbf{x}}) = \mathbb{E}[\underline{\mathbf{x}}\underline{\mathbf{x}}^T]$, the joint Gaussian distribution has maximum differential entropy [18, Thm. 8.6.5], i.e.,

$$h(\underline{\mathbf{x}}) \leq \frac{1}{2} \log \left((2\pi e)^L \det \mathbf{R}(\underline{\mathbf{x}}) \right). \quad (38)$$

Using Hadamard's inequality [18, Thm. 17.9.2] and Jensen's inequality [18, Sec. 2.6] we can further upper-bound (38) as [1]

$$h(\underline{\mathbf{x}}) \leq \frac{L}{2} \log \left(\frac{2\pi e}{L} \text{Tr}[\mathbf{R}(\underline{\mathbf{x}})] \right) \quad (39)$$

$$= \frac{L}{2} \log \left(\frac{2\pi e}{L} \mathbb{E}[\|\underline{\mathbf{x}}\|^2] \right). \quad (40)$$

A.II. Polar Coordinate System

The differential entropy of a complex random variable \mathbf{x} can be computed in polar coordinates as [19, Lemma 6.16]

$$h(\mathbf{x}) = h(|\mathbf{x}|, \angle \mathbf{x}) + \mathbb{E}[\log |\mathbf{x}|]. \quad (41)$$

Here, $\angle \mathbf{x}$ denotes the phase of \mathbf{x} . Furthermore, [19, Lemma 6.15]

$$h(|\mathbf{x}|^2) = h(|\mathbf{x}|) + \mathbb{E}[\log |\mathbf{x}|] + \log 2. \quad (42)$$

Using (41) and (42), we can upper-bound $h(\mathbf{x})$ as

$$h(\mathbf{x}) = h(|\mathbf{x}|) + h(\underline{\mathbf{z}}||\mathbf{x}|) + \mathbb{E}[\log |\mathbf{x}|] \quad (43)$$

$$= h(|\mathbf{x}|^2) - \log 2 + h(\underline{\mathbf{z}}||\mathbf{x}|) \quad (44)$$

$$\leq h(|\mathbf{x}|^2) + \log \pi. \quad (45)$$

In the last step, we used that $h(\underline{\mathbf{z}}||\mathbf{x}|) \leq h(\underline{\mathbf{z}}) \leq \log 2\pi$. Extending this inequality to L -dimensional complex random vectors, we obtain

$$h(\mathbf{x}) \leq L \log \pi + h(\llbracket \mathbf{x} \rrbracket^2). \quad (46)$$

APPENDIX B

PROOF OF LEMMA 1

Since the nonlinear step is memoryless (see (2)), \mathbf{a}_1^{nl} remains i.i.d.. Also, since the phase $\angle \mathbf{a}_{0,r}$ of $\mathbf{a}_{0,r}$, $r = 0, \dots, L-1$, is distributed uniformly over the interval $[-\pi, \pi)$ and is independent of $|\mathbf{a}_{0,r}|$, the random variable $\angle \mathbf{a}_{1,r}^{\text{nl}}$, which is equal to $\angle \mathbf{a}_{0,r} + \Delta_z \gamma |\mathbf{a}_{0,r}|^2$, is also uniformly distributed over $[-\pi, \pi)$ and independent of $|\mathbf{a}_{1,r}^{\text{nl}}|$. Since $|\mathbf{a}_{1,r}^{\text{nl}}| = |\mathbf{a}_{0,r}|$, we conclude that $\mathbf{a}_1^{\text{nl}} \sim \mathcal{CN}(\mathbf{0}, P\mathbf{I}_L)$. Furthermore, since \mathbf{U}^{li} is a unitary matrix, $\mathbf{a}_1^{\text{li}} \sim \mathcal{CN}(\mathbf{0}, P\mathbf{I}_L)$. Therefore, after noise addition, we conclude that $\mathbf{a}_1 \sim \mathcal{CN}(\mathbf{0}, (P + \sigma_n^2)\mathbf{I}_L)$. Repeating the same calculations K times and using (9), we obtain (14).

APPENDIX C

PROOF OF LEMMA 2

It follows by a conditional version of (46) that

$$\begin{aligned} & h(\mathbf{y}|\mathbf{x} = \underline{x}) \\ & \leq L \log \pi + h(\llbracket \mathbf{y} \rrbracket^2 | \mathbf{x} = \underline{x}) \end{aligned} \quad (47)$$

$$= L \log \pi + h(\llbracket \mathbf{y} \rrbracket^2 - \mathbb{E}_{\mathbf{y}}[\llbracket \mathbf{y} \rrbracket^2 | \mathbf{x} = \underline{x}] | \mathbf{x} = \underline{x}). \quad (48)$$

The last step follows because differential entropy is invariant to translations [18, Thm. 8.6.3].

For every $\underline{x} \in \mathbb{C}^L$, we define the random vector $\underline{\mathbf{w}}(\underline{x}) = \llbracket \underline{\mathbf{y}} \rrbracket^2 - \mathbb{E}[\llbracket \underline{\mathbf{y}} \rrbracket^2 | \underline{\mathbf{x}} = \underline{x}]$. Since $\underline{\mathbf{w}}(\underline{x})$ has real entries, we can use (40) to obtain

$$h(\underline{\mathbf{y}} | \underline{\mathbf{x}} = \underline{x}) \leq \frac{L}{2} \log \left(\frac{2\pi^3 e}{L} \mathbb{E}_{\underline{\mathbf{y}}} [\|\underline{\mathbf{w}}(\underline{x})\|^2 | \underline{\mathbf{x}} = \underline{x}] \right). \quad (49)$$

Averaging both sides of (49) with respect to $\underline{\mathbf{x}}$, we obtain

$$h(\underline{\mathbf{y}} | \underline{\mathbf{x}}) = \frac{L}{2} \mathbb{E}_{\underline{\mathbf{x}}} \left[\log \left(\frac{2\pi^3 e}{L} \mathbb{E}_{\underline{\mathbf{y}}} [\|\underline{\mathbf{w}}(\underline{\mathbf{x}})\|^2] \right) \right] \quad (50)$$

$$\leq \frac{L}{2} \log \left(\frac{2\pi^3 e}{L} \mathbb{E}_{\underline{\mathbf{x}}} [\mathbb{E}_{\underline{\mathbf{y}}} [\|\underline{\mathbf{w}}(\underline{\mathbf{x}})\|^2]] \right). \quad (51)$$

Here, the last inequality follows from Jensen's inequality. To conclude the proof, we note that

$$\mathbb{E} [\|\underline{\mathbf{w}}(\underline{\mathbf{x}})\|^2] = \mathbb{E} \left[\sum_{i=0}^{L-1} \left(|\mathbf{y}_i|^2 - \mathbb{E} [|\mathbf{y}_i|^2 | \underline{\mathbf{x}}] \right)^2 \right] \quad (52)$$

$$= \sum_{i=0}^{L-1} \mathbb{E} [|\mathbf{y}_i|^4] - \mathbb{E}_{\underline{\mathbf{x}}} [\mathbb{E}_{\underline{\mathbf{y}}} [|\mathbf{y}_i|^2 | \underline{\mathbf{x}}]^2] \quad (53)$$

$$= \kappa(\underline{\mathbf{y}} | \underline{\mathbf{x}}). \quad (54)$$

APPENDIX D

PROOF OF (29)

To prove (29), we start by noting that (see Fig. 1)

$$\mathbb{E} [|\mathbf{a}_{K,r}|^2 | \underline{\mathbf{a}}_0] = \mathbb{E} [|\mathbf{a}_{K,r}^{\text{li}}|^2 | \underline{\mathbf{a}}_0] + \sigma_n^2 \quad (55)$$

$$\geq \mathbb{E} [|\mathbf{a}_{K,r}^{\text{li}}|^2 | \underline{\mathbf{a}}_0]. \quad (56)$$

In Appendix D.I, we prove that for every $2 \leq k \leq K$,

$$\mathbb{E} [|\mathbf{a}_{k,r}^{\text{li}}|^2 | \underline{\mathbf{a}}_0] \geq \mathbb{E} [|\mathbf{a}_{k-1,r}|^2 | \underline{\mathbf{a}}_0] |\alpha_{0,0}|^2 - \frac{D}{K} \quad (57)$$

where $\alpha_{0,0}$ and D are defined in (24) and (26), respectively. Using (56) and (57) $K-1$ times, we obtain (recall that $\mathbf{a}_{1,r}^{\text{li}}$ is a deterministic function of \mathbf{a}_0)

$$\mathbb{E}\left[|\mathbf{a}_{K,r}|^2 \middle| \mathbf{a}_0\right] \geq |\mathbf{a}_{1,r}^{\text{li}}|^2 |\alpha_{0,0}|^{2(K-1)} - \frac{D}{K} \sum_{j=0}^{K-2} |\alpha_{0,0}|^{2j}. \quad (58)$$

Furthermore, since $|\alpha_{0,0}| \leq 1$ (see (24)), we conclude that

$$\mathbb{E}\left[|\mathbf{a}_{K,r}|^2 \middle| \mathbf{a}_0\right] \geq |\mathbf{a}_{1,r}^{\text{li}}|^2 |\alpha_{0,0}|^{2(K-1)} - D. \quad (59)$$

Squaring both sides of (59), we get

$$\left(\mathbb{E}\left[|\mathbf{a}_{K,r}|^2 \middle| \mathbf{a}_0\right]\right)^2 \geq \left(|\mathbf{a}_{1,r}^{\text{li}}|^2 |\alpha_{0,0}|^{2(K-1)} - D\right)^2 \times \mathbb{1}\left(|\mathbf{a}_{1,r}^{\text{li}}|^2 > D/|\alpha_{0,0}|^{2(K-1)}\right). \quad (60)$$

We know from the proof of Lemma 1 that $\mathbf{a}_1^{\text{li}} \sim \mathcal{CN}(\underline{0}, P\mathbf{I}_L)$. Consequently, $|\mathbf{a}_{1,r}^{\text{li}}|^2 \sim \text{Exp}(1/P)$.

Averaging both sides of (60) with respect to \mathbf{a}_0 , we obtain

$$\mathbb{E}\left[\left(\mathbb{E}\left[|\mathbf{a}_{K,r}|^2 \middle| \mathbf{a}_0\right]\right)^2\right] \geq \frac{1}{P} \int_{D/|\alpha_{0,0}|^{2(K-1)}}^{\infty} \left(a|\alpha_{0,0}|^{2(K-1)} - D\right)^2 \exp\left(-\frac{a}{P}\right) da \quad (61)$$

$$= 2P^2 |\alpha_{0,0}|^{4(K-1)} \exp\left(\frac{-D}{P|\alpha_{0,0}|^{2(K-1)}}\right) \quad (62)$$

$$= 2P^2 \zeta(P, K) \quad (63)$$

where in the last step we used (27).

D.I. Proof of (57)

Let $\tilde{\mathbf{a}}_k^{\text{nl}} = \mathbf{F}\mathbf{a}_k^{\text{nl}}$. By definition of the DFT operator \mathbf{F} in (5), we have

$$\tilde{\mathbf{a}}_{k,n}^{\text{nl}} = \frac{1}{\sqrt{L}} \sum_{m=0}^{L-1} \mathbf{a}_{k,m}^{\text{nl}} e^{-j\frac{2\pi}{L}mn}. \quad (64)$$

Furthermore, it follows from (5) and (6) that

$$\mathbf{a}_{k,r}^{\text{li}} = \frac{1}{\sqrt{L}} \sum_{n=0}^{L-1} \tilde{\mathbf{a}}_{k,n}^{\text{nl}} e^{j\frac{2\pi}{L}nr} e^{j\Delta_z f^{\text{li}}(l)} \quad (65)$$

$$= \frac{1}{L} \sum_{n=0}^{L-1} \sum_{m=0}^{L-1} \mathbf{a}_{k,m}^{\text{nl}} e^{j\frac{2\pi}{L}n(r-m)} e^{j\Delta_z f^{\text{li}}(l)} \quad (66)$$

$$= \frac{1}{L} \sum_{m=0}^{L-1} \mathbf{a}_{k,m}^{\text{nl}} \sum_{n=0}^{L-1} e^{j\frac{2\pi}{L}n(r-m)} e^{j\Delta_z f^{\text{li}}(l)} \quad (67)$$

$$= \sum_{m=0}^{L-1} \mathbf{a}_{k,m}^{\text{nl}} \alpha_{r,m} \quad (68)$$

where

$$\alpha_{r,m} = \frac{1}{L} \sum_{n=0}^{L-1} e^{j\frac{2\pi}{L}n(r-m)} e^{j\Delta_z f^{\text{li}}(l)}. \quad (69)$$

By squaring both sides of (68), we obtain

$$|\mathbf{a}_{k,r}^{\text{li}}|^2 = \sum_{m=0}^{L-1} |\mathbf{a}_{k,m}^{\text{nl}}|^2 |\alpha_{r,m}|^2 + \sum_{\substack{m,n \\ m \neq n}} \mathbf{a}_{k,m}^{\text{nl}} (\mathbf{a}_{k,n}^{\text{nl}})^* \alpha_{r,m} \alpha_{r,n}^*. \quad (70)$$

Taking the expectation of both sides of (70) and using triangle inequality, we get

$$\mathbb{E} [|\mathbf{a}_{k,r}^{\text{li}}|^2 | \underline{\mathbf{a}}_0] = \sum_m \mathbb{E} [|\mathbf{a}_{k,m}^{\text{nl}}|^2] |\alpha_{r,m}|^2 + \sum_{\substack{m,n \\ m \neq n}} \mathbb{E} [\mathbf{a}_{k,m}^{\text{nl}} (\mathbf{a}_{k,n}^{\text{nl}})^*] \alpha_{r,m} \alpha_{r,n}^* \quad (71)$$

$$\geq \mathbb{E} [|\mathbf{a}_{k,r}^{\text{nl}}|^2 | \underline{\mathbf{a}}_0] |\alpha_{r,r}|^2 - \left| \sum_{\substack{m,n \\ m \neq n}} \mathbb{E} [\mathbf{a}_{k,m}^{\text{nl}} (\mathbf{a}_{k,n}^{\text{nl}})^*] \alpha_{r,m} \alpha_{r,n}^* \right| \quad (72)$$

$$\geq \mathbb{E} [|\mathbf{a}_{k,r}^{\text{nl}}|^2 | \underline{\mathbf{a}}_0] |\alpha_{r,r}|^2 - \sum_{\substack{m,n \\ m \neq n}} \left| \mathbb{E} [\mathbf{a}_{k,m}^{\text{nl}} (\mathbf{a}_{k,n}^{\text{nl}})^* | \underline{\mathbf{a}}_0] \right| |\alpha_{r,m}| |\alpha_{r,n}| \quad (73)$$

$$\geq \mathbb{E} [|\mathbf{a}_{k,r}^{\text{nl}}|^2 | \underline{\mathbf{a}}_0] |\alpha_{r,r}|^2 - \max_{m,n} \left\{ \left| \mathbb{E} [\mathbf{a}_{k,m}^{\text{nl}} (\mathbf{a}_{k,n}^{\text{nl}})^* | \underline{\mathbf{a}}_0] \right| \right\} \sum_{\substack{m,n \\ m \neq n}} |\alpha_{r,m}| |\alpha_{r,n}|. \quad (74)$$

We note that since $\alpha_{r,n}$ depends on (r, m) only through $r - m$ (see (69)), for every $r = 0, \dots, L - 1$, we have

$$|\alpha_{r,r}|^2 = |\alpha_{0,0}|^2 \quad (75)$$

and

$$\sum_{\substack{m,n \\ m \neq n}} |\alpha_{r,n}| |\alpha_{r,m}| = \sum_{\substack{m,n \\ m \neq n}} |\alpha_{0,n}| |\alpha_{0,m}| \quad (76)$$

$$= \left(\sum_{m=0}^{L-1} |\alpha_{0,m}| \right)^2 - \sum_{m=0}^{L-1} |\alpha_{0,m}|^2 = \alpha \quad (77)$$

where in the last step we used (25). In Appendix D.II, we prove that for all $2 \leq k \leq K$ and all $0 \leq m, n \leq L-1$,

$$\left| \mathbb{E} \left[\mathbf{a}_{k,m}^{\text{nl}} \left(\mathbf{a}_{k,n}^{\text{nl}} \right)^* \middle| \mathbf{a}_0 \right] \right| \leq \frac{K^3}{2e\gamma^2 P_n Z^2} + \frac{K}{eZ\gamma} + \frac{P_n}{2eK} = \frac{D}{\alpha K} \quad (78)$$

where in the last step we used (26). Since $|\mathbf{a}_{k,r}^{\text{nl}}| = |\mathbf{a}_{k-1,r}|$, we obtain (57) by substituting (75), (76), and (78) into (74).

D.II. Proof of (78)

For every $0 \leq m, n \leq L-1$ and $1 \leq k \leq K-1$, we have

$$\left| \mathbb{E} \left[\mathbf{a}_{k+1,m}^{\text{nl}} \left(\mathbf{a}_{k+1,n}^{\text{nl}} \right)^* \middle| \mathbf{a}_0 \right] \right| = \left| \mathbb{E} \left[\mathbb{E}_{\mathbf{n}_{k,m}} \left[\mathbf{a}_{k+1,m}^{\text{nl}} \right] \mathbb{E}_{\mathbf{n}_{k,n}} \left[\left(\mathbf{a}_{k+1,n}^{\text{nl}} \right)^* \right] \middle| \mathbf{a}_0 \right] \right| \quad (79)$$

$$\leq \left(\max_{\mathbf{a}_{k,m}^{\text{li}}} \left\{ \left| \mathbb{E}_{\mathbf{n}_{k,m}} \left[\mathbf{a}_{k+1,m}^{\text{nl}} \right] \right| \right\} \right) \left(\max_{\mathbf{a}_{k,n}^{\text{li}}} \left\{ \left| \mathbb{E}_{\mathbf{n}_{k,n}} \left[\left(\mathbf{a}_{k+1,n}^{\text{nl}} \right)^* \right] \right| \right\} \right). \quad (80)$$

The last step follows because

$$\mathbf{a}_{k+1,m}^{\text{nl}} = \left(\mathbf{a}_{k,m}^{\text{nl}} + \mathbf{n}_{k,m} \right) e^{j\gamma\Delta_z |\mathbf{a}_{k,m}^{\text{li}} + \mathbf{n}_{k,m}|^2}. \quad (81)$$

Using triangle inequality, we get that

$$\begin{aligned} \max_{\mathbf{a}_{k,m}^{\text{li}}} \left\{ \left| \mathbb{E}_{\mathbf{n}_{k,m}} \left[\mathbf{a}_{k+1,m}^{\text{nl}} \right] \right| \right\} &\leq \max_{\mathbf{a}_{k,m}^{\text{li}}} \left\{ \left| \mathbf{a}_{k,m}^{\text{li}} \right| \cdot \left| \mathbb{E}_{\mathbf{n}_{k,m}} \left[e^{j\Delta_z \gamma (|\mathbf{a}_{k,m}^{\text{li}} + \mathbf{n}_{k,m}|^2)} \right] \right| \right\} \\ &\quad + \max_{\mathbf{a}_{k,m}^{\text{li}}} \left\{ \left| \mathbb{E}_{\mathbf{n}_{k,m}} \left[\mathbf{n}_{k,m} e^{j\Delta_z \gamma (|\mathbf{a}_{k,m}^{\text{li}} + \mathbf{n}_{k,m}|^2)} \right] \right| \right\}. \end{aligned} \quad (82)$$

We first compute the first term on the RHS of (82). Note that given $\mathbf{a}_{k,m}^{\text{li}}$, the random variable

$$\mathbf{w} = \frac{2|\mathbf{a}_{k,m}^{\text{li}} + \mathbf{n}_{k,m}|^2}{\sigma_n^2} \quad (83)$$

which is proportional to the argument of the exponential term on the RHS of (82), follows a noncentral chi-squared distribution with two degrees of freedom and noncentrality parameter $\lambda = 2|\mathbf{a}_{k,m}^{\text{li}}|^2/\sigma_n^2$. Let $M_{\mathbf{w}}(t)$ be the moment generating function of \mathbf{w} . We have [20, Chap. 7]

$$M_{\mathbf{w}}(t) = \mathbb{E}[e^{t\mathbf{w}}] \quad (84)$$

$$= \frac{1}{1-2t} \exp\left(\frac{\lambda t}{1-2t}\right). \quad (85)$$

Therefore,

$$|\mathbf{a}_{k,m}^{\text{li}}| \cdot \left| \mathbb{E}_{\mathbf{n}_{k,m}} \left[e^{j\Delta_z \gamma (|\mathbf{a}_{k,m}^{\text{li}} + \mathbf{n}_{k,m}|^2)} \right] \right| = |\mathbf{a}_{k,m}^{\text{li}}| \left| M_{\mathbf{w}}\left(\frac{j\Delta_z \gamma \sigma_n^2}{2}\right) \right| \quad (86)$$

$$= \frac{|\mathbf{a}_{k,m}^{\text{li}}|}{\sqrt{1 + \sigma_n^4 \gamma^2 \Delta_z^2}} \exp\left(-\frac{\sigma_n^2 \gamma^2 \Delta_z^2}{1 + \sigma_n^4 \gamma^2 \Delta_z^2} |\mathbf{a}_{k,m}^{\text{li}}|^2\right). \quad (87)$$

Using in (87) that

$$\max_{x>0} \{x e^{-ax^2}\} = \frac{1}{\sqrt{2ea}}, \quad a > 0 \quad (88)$$

we obtain

$$\max_{\mathbf{a}_{k,m}^{\text{li}}} \left\{ |\mathbf{a}_{k,m}^{\text{li}}| \cdot \left| \mathbb{E}_{\mathbf{n}_{k,m}} \left[e^{j\Delta_z \gamma (|\mathbf{a}_{k,m}^{\text{li}} + \mathbf{n}_{k,m}|^2)} \right] \right| \right\} = \frac{1}{\sqrt{2e\sigma_n \gamma \Delta_z}}. \quad (89)$$

The evaluation of the second term on the RHS of (82) requires same care. Indeed, although the phase of $\mathbf{n}_{k,m}$ is uniformly distributed, the phase of $\mathbf{n}_{k,m} \exp(j\Delta_z \gamma (|\mathbf{a}_{k,m}^{\text{li}} + \mathbf{n}_{k,m}|^2))$ is not uniform since $|\mathbf{a}_{k,m}^{\text{li}} + \mathbf{n}_{k,m}|^2$ depends on $\angle \mathbf{n}_{k,m}$. To evaluate this term, we first note that

$$\left| \mathbb{E}_{\mathbf{n}_{k,m}} \left[\mathbf{n}_{k,m} e^{j\Delta_z \gamma (|\mathbf{a}_{k,m}^{\text{li}} + \mathbf{n}_{k,m}|^2)} \right] \right| = \left| \mathbb{E}_{\mathbf{n}_{k,m}} \left[\mathbf{n}_{k,m} e^{j\Delta_z \gamma (|\mathbf{n}_{k,m}|^2 + 2\mathcal{R}(\mathbf{a}_{k,m}^{\text{li}} \mathbf{n}_{k,m}^*))} \right] \right|. \quad (90)$$

We then separate the real and imaginary parts of $\mathbf{n}_{k,m}$ on the RHS of (90) to obtain

$$\begin{aligned} & \mathbb{E}_{\mathbf{n}_{k,m}} \left[\mathbf{n}_{k,m} e^{j\Delta_z \gamma (|\mathbf{n}_{k,m}|^2 + 2\mathcal{R}(\mathbf{a}_{k,m}^{\text{li}} \mathbf{n}_{k,m}^*))} \right] \\ &= \mathbb{E}_{\mathcal{R}(\mathbf{n}_{k,m})} \left[\mathcal{R}(\mathbf{n}_{k,m}) e^{j\gamma \Delta_z (\mathcal{R}(\mathbf{n}_{k,m})^2 + 2\mathcal{R}(\mathbf{a}_{k,m}^{\text{li}}) \mathcal{R}(\mathbf{n}_{k,m}))} \right] \\ & \quad \times \mathbb{E}_{\mathcal{I}(\mathbf{n}_{k,m})} \left[e^{j\gamma \Delta_z (\mathcal{I}(\mathbf{n}_{k,m})^2 + 2\mathcal{I}(\mathbf{a}_{k,m}^{\text{li}}) \mathcal{I}(\mathbf{n}_{k,m}))} \right] \end{aligned}$$

$$\begin{aligned}
& + jE_{\mathcal{R}(\mathbf{n}_{k,m})} \left[e^{j\gamma\Delta_z \left(\mathcal{R}(\mathbf{n}_{k,m})^2 + 2\mathcal{R}(\mathbf{a}_{k,m}^{\text{li}}) \mathcal{R}(\mathbf{n}_{k,m}) \right)} \right] \\
& \times E_{\mathcal{I}(\mathbf{n}_{k,m})} \left[\mathcal{I}(\mathbf{n}_{k,m}) e^{j\gamma\Delta_z \left(\mathcal{I}(\mathbf{n}_{k,m})^2 + 2\mathcal{I}(\mathbf{a}_{k,m}^{\text{li}}) \mathcal{I}(\mathbf{n}_{k,m}) \right)} \right].
\end{aligned} \tag{91}$$

One can calculate the expectations on the RHS of (91) in closed form by writing them as integrals involving the Gaussian probability density function and by using the following equalities:

$$\int_{-\infty}^{\infty} x e^{ax^2+bx} dx = \frac{b\sqrt{\pi}}{2(-a)^{3/2}} e^{\frac{-b^2}{4a}}, \quad \mathcal{R}(a) < 0 \tag{92}$$

and

$$\int_{-\infty}^{\infty} e^{ax^2+bx} dx = e^{\frac{-b^2}{4a}} \sqrt{\frac{\pi}{-a}}, \quad \mathcal{R}(a) < 0. \tag{93}$$

Through these steps, one gets

$$E_{\mathbf{n}_{k,m}} \left[\mathbf{n}_{k,m} e^{j\Delta_z \gamma (|\mathbf{n}_{k,m}|^2 + 2\mathcal{R}(\mathbf{a}_{k,m}^{\text{li}} \mathbf{n}_{k,m}^*))} \right] = \frac{j\sigma_n^2 \gamma \Delta_z}{(1 - j\sigma_n^2 \gamma \Delta_z)^2} \mathbf{a}_{k,m}^{\text{li}} \exp \left(-\frac{(\sigma_n \gamma \Delta_z |\mathbf{a}_{k,m}^{\text{li}}|)^2}{1 - j\sigma_n^2 \gamma \Delta_z} \right). \tag{94}$$

Substituting (94) into (90), we obtain

$$\left| E_{\mathbf{n}_{k,m}} \left[\mathbf{n}_{k,m} e^{j\Delta_z \gamma (|\mathbf{a}_{k,m}^{\text{li}} + \mathbf{n}_{k,m}|^2)} \right] \right| = \frac{\sigma_n^2 \gamma \Delta_z}{1 + \sigma_n^4 \gamma^2 \Delta_z^2} |\mathbf{a}_{k,m}^{\text{li}}| \exp \left(-\frac{(\sigma_n \gamma \Delta_z |\mathbf{a}_{k,m}^{\text{li}}|)^2}{1 + \sigma_n^4 \gamma^2 \Delta_z^2} \right). \tag{95}$$

Using (88) in (95) we conclude that

$$\max_{\mathbf{a}_k^{\text{li}}} \left\{ \left| E_{\mathbf{n}_{k,m}} \left[\mathbf{n}_{k,m} e^{j\Delta_z \gamma (|\mathbf{a}_{k,m}^{\text{li}} + \mathbf{n}_{k,m}|^2)} \right] \right| \right\} = \frac{\sigma_n}{\sqrt{2e(1 + \sigma_n^4 \gamma^2 \Delta_z^2)}}. \tag{96}$$

Furthermore, substituting (89) and (96) into (82), we obtain

$$\max_{\mathbf{a}_k^{\text{li}}} \left\{ \left| E_{\mathbf{n}_{k,m}} \left[\mathbf{a}_{k+1,m}^{\text{nl}} \right] \right| \right\} \leq \frac{1}{\sigma_n \gamma \Delta_z \sqrt{2e}} + \frac{\sigma_n}{\sqrt{2e(1 + \sigma_n^4 \gamma^2 \Delta_z^2)}} \tag{97}$$

$$\leq \frac{K\sqrt{K}}{\gamma Z \sqrt{2eP_n}} + \frac{\sqrt{P_n}}{\sqrt{2eK}}. \tag{98}$$

Proceeding analogously, one can show that the same bound holds for the second term on the RHS of (80). Namely,

$$\max_{\mathbf{a}_{k,m}^{\text{li}}} \left\{ \left| E_{\mathbf{n}_{k,n}} \left[(\mathbf{a}_{k+1,n}^{\text{nl}})^* \right] \right| \right\} \leq \frac{K\sqrt{K}}{\gamma Z \sqrt{2eP_n}} + \frac{\sqrt{P_n}}{\sqrt{2eK}}. \tag{99}$$

Substituting (98) and (99) into (80), we obtain after simple algebraic manipulations

$$\left| \mathbb{E} \left[\mathbf{a}_{k+1,m}^{\text{nl}} \left(\mathbf{a}_{k+1,n}^{\text{nl}} \right)^* \middle| \mathbf{a}_0 \right] \right| \leq \frac{K^3}{2e\gamma^2 P_n Z^2} + \frac{K}{eZ\gamma} + \frac{P_n}{2eK}. \quad (100)$$

We note that the steps just performed are not applicable to the first segment ($k = 0$) of the SSF channel as there is no noise.

APPENDIX E

PROOF OF THEOREM 3

Comparing (28) and (34), one sees that to establish (34), it is sufficient to show that

$$\zeta(P)P^2 \geq P^2 - \frac{2C_1}{K}P^2 - PG \quad (101)$$

or, equivalently, that

$$1 - \zeta(P) \leq \frac{2C_1}{K} + \frac{G}{P}. \quad (102)$$

It follows from (27) that

$$\zeta(P) = |\alpha_{0,0}|^{4(K-1)} \exp \left(-\frac{D}{P|\alpha_{0,0}|^{2(K-1)}} \right) \quad (103)$$

$$\geq |\alpha_{0,0}|^{4(K-1)} \left(1 - \frac{D}{P|\alpha_{0,0}|^{2(K-1)}} \right) \quad (104)$$

where in the last step we used that $e^{-x} \geq 1 - x$ for all $x \in \mathbb{R}$. In Appendix E.I, we prove that for every $K > \frac{\beta_2 Z \pi^2}{2\sqrt{2}\Delta_t^2}$ we have

$$|\alpha_{0,0}|^2 \geq 1 - \frac{C_1}{K^2}. \quad (105)$$

Substituting (105) into (104), and using that $|\alpha_{0,0}| \leq 1$ (see the definition in (24)), we obtain

$$\zeta(P) = |\alpha_{0,0}|^{4(K-1)} - |\alpha_{0,0}|^{2(K-1)} \frac{D}{P} \quad (106)$$

$$\geq \left(1 - \frac{C_1}{K^2} \right)^{2K} - \frac{D}{P}. \quad (107)$$

Since the function $f(C_1) = (1 - C_1/K^2)^{2K}$ is convex in the interval $[0, K^2]$, for every $C_1 \leq K^2$ or equivalently every $K \geq \sqrt{C_1}$, we have

$$f(C_1) \geq f(0) + f'(0)C_1. \quad (108)$$

It follows from (108) that

$$\left(1 - \frac{C_1}{K^2}\right)^{2K} \geq 1 - \frac{2C_1}{K}. \quad (109)$$

Substituting (109) into (107) we obtain

$$1 - \zeta(P) \leq \frac{2C_1}{K} + \frac{D}{P}. \quad (110)$$

To conclude the proof, we show that $D \leq G$ when $K > \frac{\beta_2 Z \pi^2}{2\sqrt{2}\Delta_t^2}$. Comparing (26) and (33), we see that this is equivalent to showing that

$$\alpha \leq \frac{C_2}{K} \left(\frac{C_2}{4K} + 1 \right) \quad (111)$$

where α and C_2 are defined in (25) and (32), respectively. To prove (111), we show in Appendix E.II that if $K > \frac{\beta_2 Z \pi^2}{2\sqrt{2}\Delta_t^2}$ then

$$|\alpha_{0,m}| \leq \frac{C_2}{2KL} \quad (112)$$

for every $1 \leq m \leq L-1$. Furthermore, by the definition of α in (25) we have

$$\alpha = \left(\sum_{m=0}^{L-1} |\alpha_{0,m}| \right)^2 - \sum_{m=0}^{L-1} |\alpha_{0,m}|^2 \quad (113)$$

$$\leq \left(\sum_{m=0}^{L-1} |\alpha_{0,m}| \right)^2 - |\alpha_{0,0}|^2 \quad (114)$$

$$= \left(|\alpha_{0,0}| + \sum_{n=1}^{L-1} |\alpha_{0,n}| \right)^2 - |\alpha_{0,0}|^2. \quad (115)$$

Using (112) in (115), we obtain

$$\alpha \leq \left(|\alpha_{0,0}| + \frac{C_2}{2K} \right)^2 - |\alpha_{0,0}|^2 \quad (116)$$

$$= \frac{C_2^2}{4K^2} + |\alpha_{0,0}| \frac{C_2}{K}. \quad (117)$$

We obtain (111) from (117) by using that $|\alpha_{0,0}| \leq 1$.

E.I. Proof of (105)

Note that

$$|\alpha_{0,0}|^2 = \frac{1}{L^2} \left| \sum_{l=0}^{L-1} e^{-j \frac{Z}{K} f^{\text{li}}(l)} \right|^2 \quad (118)$$

$$\geq \frac{1}{L^2} \left(\sum_{l=0}^{L-1} \cos \left(\frac{Z}{K} f^{\text{li}}(l) \right) \right)^2. \quad (119)$$

Using the inequality $\cos(x) \geq 1 - x^2/2$, we obtain

$$\sum_{l=0}^{L-1} \cos \left(\frac{Z}{K} f^{\text{li}}(l) \right) \geq \sum_{l=0}^{L-1} \left(1 - \frac{Z^2 (f^{\text{li}}(l))^2}{2K^2} \right). \quad (120)$$

One can verify that for every K such that

$$K > \frac{Z}{\sqrt{2}} \max_l f^{\text{li}}(l) \geq \frac{\beta_2 Z \pi^2}{2\sqrt{2}\Delta_t^2} \quad (121)$$

the RHS of (120) is positive and the inequality in (120) holds also when we square both sides. Thus, if (121) holds, we have that

$$|\alpha_{0,0}|^2 \geq \frac{1}{L^2} \left(L - \frac{Z^2}{2K^2} \sum_{l=0}^{L-1} (f^{\text{li}}(l))^2 \right)^2 \quad (122)$$

$$= \left(1 - \frac{C_1}{2K^2} \right)^2 \quad (123)$$

$$\geq 1 - \frac{C_1}{K^2}. \quad (124)$$

Here, (123) follows from (31) and in (124) we used that $(1 - x)^2 \geq 1 - 2x$.

E.II. Proof of (112)

If $m \neq 0$, we have (see (24))

$$\alpha_{0,m} = \frac{1}{L} \sum_{l=0}^{L-1} e^{-j \frac{2\pi}{L} lm} e^{-j \frac{Z}{K} f^{\text{li}}(l)} \quad (125)$$

$$= \frac{1}{L} \sum_{l=0}^{L-1} e^{-j \frac{2\pi}{L} lm} \left(e^{-j \frac{Z}{K} f^{\text{li}}(l)} - 1 \right). \quad (126)$$

Using triangle inequality,

$$|\alpha_{0,m}| \leq \frac{1}{L} \sum_{l=0}^{L-1} \left| e^{-j\frac{2\pi}{L}lm} \right| \left| e^{-j\frac{Z}{K}f^{\text{li}}(l)} - 1 \right| \quad (127)$$

$$= \frac{1}{L} \sum_{l=0}^{L-1} \sqrt{\left(1 - \cos\left(\frac{Z}{K}f^{\text{li}}(l)\right)\right)^2 + \left(\sin\left(\frac{Z}{K}f^{\text{li}}(l)\right)\right)^2}. \quad (128)$$

Furthermore, using that $\cos(x) \geq 1 - x^2/2$ for every x and $\sin(x) \leq x$ for every $x \geq 0$, we get

$$|\alpha_{r,m}| \leq \frac{1}{L} \sum_{l=0}^{L-1} \sqrt{\left(\frac{Z}{\sqrt{2}K}f^{\text{li}}(l)\right)^4 + \left(\frac{Z}{K}f^{\text{li}}(l)\right)^2}. \quad (129)$$

For every $K > \frac{\beta_2 Z \pi^2}{2\sqrt{2}\Delta_t^2}$, we have that $\frac{Z}{\sqrt{2}K}f^{\text{li}}(l) \leq 1$. Thus,

$$\left(\frac{Z}{\sqrt{2}K}f^{\text{li}}(l)\right)^4 \leq \left(\frac{Z}{\sqrt{2}K}f^{\text{li}}(l)\right)^2. \quad (130)$$

Let C_2 be defined as in (32). Substituting (130) into (129), we obtain

$$|\alpha_{0,m}| \leq \frac{C_2}{2KL}. \quad (131)$$

ACKNOWLEDGMENT

The authors would like to thank L. Barletta, N. V. Irukulapati, G. Kramer and M. I. Yousefi for helpful discussions and comments on an early version of this paper.

REFERENCES

- [1] G. Kramer, M. I. Yousefi, and F. R. Kschischang, "Upper bound on the capacity of a cascade of nonlinear and noisy channels," in *IEEE Info. Theory Workshop (ITW)*, Apr.–May 2015.
- [2] M. I. Yousefi, G. Kramer, and F. R. Kschischang, "Upper bound on the capacity of the nonlinear Schrödinger channel," in *Can. Workshop on Inf. Theory (CWIT)*, Jul. 2015.
- [3] M. I. Yousefi and F. R. Kschischang, "On the per-sample capacity of nondispersive optical fibers," *IEEE Trans. Inform. Theory*, vol. 57, no. 11, pp. 7522–7541, Nov. 2011.
- [4] K. S. Turitsyn, S. A. Derevyanko, I. Yurkevich, and S. K. Turitsyn, "Information capacity of optical fiber channels with zero average dispersion," *Phys. Rev. Lett.*, vol. 91, no. 20, p. 203901, Nov. 2003.
- [5] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature*, vol. 411, no. 6841, pp. 1027–1030, 2001.

- [6] A. D. Ellis, J. Zhao, and D. Cotter, “Approaching the non-linear Shannon limit,” *J. Lightw. Technol.*, vol. 28, no. 4, pp. 423–433, Feb. 2010.
- [7] A. Meccozzi and R.-J. Essiambre, “Nonlinear Shannon limit in pseudolinear coherent systems,” *J. Lightw. Technol.*, vol. 30, no. 12, pp. 2011–2024, Jun. 2012.
- [8] N. V. Irukulapati, M. Secondini, E. Agrell, P. Johannisson, and H. Wymeersch, “Tighter lower bounds on mutual information for fiber-optic channels,” *preprint arXiv:1606.09176*, 2016.
- [9] M. Secondini, E. Forestieri, and G. Prati, “Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps,” *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3839–3852, Dec. 2013.
- [10] E. Agrell, “Conditions for a monotonic channel capacity,” *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 738–748, 2015.
- [11] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, “Capacity of a nonlinear optical channel with finite memory,” *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2862–2876, Aug. 2014.
- [12] G. P. Agrawal, *Nonlinear Fiber Optics*, 4th ed. Academic Press, Boston, 2007.
- [13] Q. Zhang and M. I. Hayee, “Symmetrized split-step Fourier scheme to control global simulation accuracy in fiber-optic communication systems,” *J. Lightw. Technol.*, vol. 26, no. 2, pp. 302–316, Jan. 2008.
- [14] A. Meccozzi, “Limits to long-haul coherent transmission set by the Kerr nonlinearity and noise of the in-line amplifiers,” *J. Lightw. Technol.*, vol. 12, no. 11, pp. 1993–2000, Nov. 1994.
- [15] A. P. T. Lau and J. M. Kahn, “Signal design and detection in presence of nonlinear phase noise,” *J. Lightw. Technol.*, vol. 25, no. 10, pp. 3008–3016, Oct. 2007.
- [16] M. Secondini and E. Forestieri, “Analytical fiber-optic channel model in the presence of cross-phase modulation,” *IEEE Photonics Technology Letters*, vol. 24, no. 22, pp. 2016–2019, 2012.
- [17] F. D. Neeser and J. L. Massey, “Proper complex random processes with applications to information theory,” *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.
- [18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, New Jersey, 2006.
- [19] A. Lapidoth and S. M. Moser, “Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels,” *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
- [20] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. McGraw-Hill Education, 2002.